GitHubUpload

Max Strul

This is my first year exam bioinformatics section

This section is the code for question 10 asking for a covid-19 variants graph

this will be uploaded to github.

```
#load libraries
library(ggplot2)
#Load data
covid_data <- read.csv("~/Desktop/Classes/FirstYearExam/FirstYearExamq10/covid19_variants.
#analyze data
head(covid_data)
```

```
        date       area area_type variant_name specimens percentage
1 2021-01-01 California     State      Omicron         1       1.67
2 2021-01-01 California     State           Mu         0       0.00
3 2021-01-01 California     State        Gamma         0       0.00
4 2021-01-01 California     State      Epsilon        29      48.33
5 2021-01-01 California     State        Other        29      48.33
6 2021-01-01 California     State        Total        60     100.00
  specimens_7d_avg percentage_7d_avg
1               NA                NA
2               NA                NA
3               NA                NA
4               NA                NA
5               NA                NA
6               NA                NA
```

```r
#make new data frame with only california data
california_data <- covid_data[covid_data$area=="California",]
#make the data for dates the "Date" data type
california_data$date <- as.Date(california_data$date)
#remove other and total data
cleaned_data_cal <- california_data[california_data$variant_name!="Other",]
cleaned_data_cal <- california_data[california_data$variant_name!="Total",]
#set the data to be from jan 2021 to may 2022
cleaned_data_cal <- cleaned_data_cal[cleaned_data_cal$date>=as.Date("2021-01-01")
                                     &cleaned_data_cal$date<=as.Date("2022-05-01"),]
#plot it
ggplot(data=cleaned_data_cal)+
  aes(x=date,y=percentage,col=variant_name)+
  geom_line()+
  theme_bw()+
  ylab("Percentage of sequenced specimens")+
  labs(caption = "Data Source: <https://www.cdph.ca.gov/>")+
  scale_x_date(date_breaks = "month",date_labels = "%b %Y")+
  theme(axis.text.x = element_text(angle = 60, hjust = 1))+
  xlab("")+ggtitle("Covid-19 Variants in California")
```

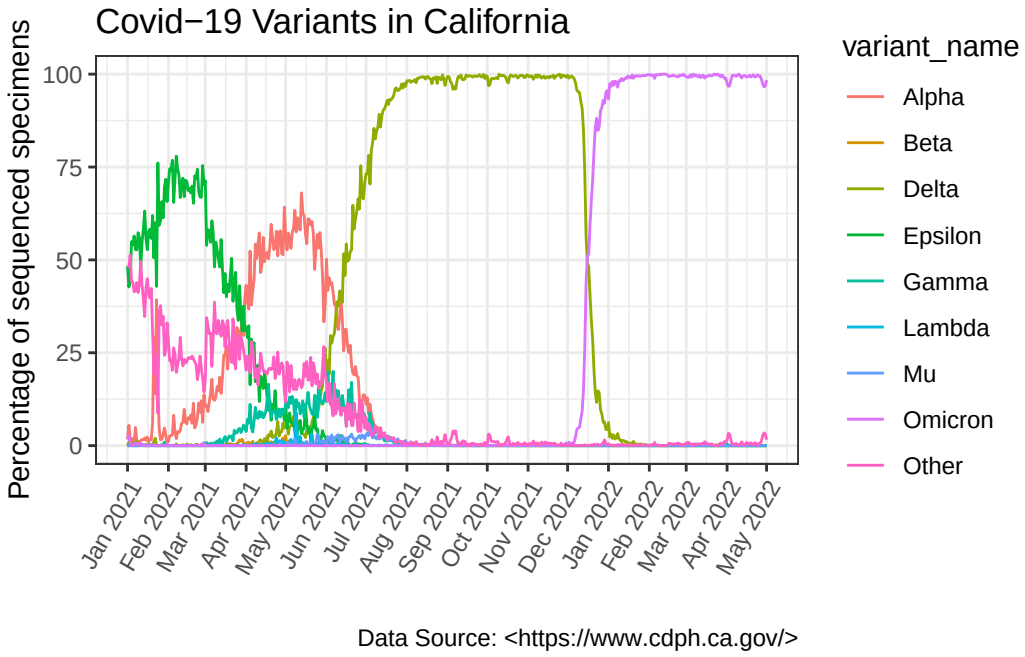

Figure 1: Plot for Q10.